

THREE TYPES OF ORAL EVALUATION  
- ANALYTIC, SEMI-ANALYTIC AND HOLISTIC -  
キーワード：音声言語、評価、信頼性

内藤 徹  
Toru Naito

## 1. はじめに

スピーチの評価は音声言語を対象とするので、文字言語と比べて後戻りできず、また時間的にも制約があり困難な点が多い。そして、その評価法についても、ANALYTIC(AL) EVALUATION (A. E.)、SEMI-ANALYTIC EVALUATION (S. E.)、HOLISTIC (IMPRESSIONISTIC) EVALUATION (H. E.)が考えられ、それぞれの関係と効果については明らかにされていない。

今回、高校生を対象とする英語弁論において、これらの評価に関するデータを得ることができた。この研究が SPEECH のみならず ORAL COMMUNICATION の評価法にも1つのヒントになれば幸いである。

## 2. 先行文献研究

### (1) SPEAKING および SPEECH の評価

測定法として、小川(1982)は、直接測定法(DIRECT MEASUREMENT)、半直接測定法(SEMI-DIRECT MEASUREMENT)、間接測定法(INDIRECT MEASUREMENT)をあげているが、本論の中では直接測定法のみを取り扱う。

また、Valette(1977)は、評価者(審査員など)は評価の信頼性を高めるために、目標言語の高い運用能力を持ち、専門的に熟練した者が、明確に定義づけられた採点基準に従って評価を行う必要があると述べている。さらに、客観的な評価をすることも大切である。Yorozuya(1980)は、ハロー効果(HALO EFFECT)によってスピーキングやスピーチの信頼性が約3%低下したと報告しており、注意を要する点であろう。

実用性からみた評価の問題点は、福井(1969)によれば、①採点に要する時間と採点のしやすさ、②実施に要する時間と場所や設備などの実施のしやすさ、の2点が考えられる。①の場合、A. E. は評価が複雑でかつ煩雑となり、時間の面でも問題になるが、評価の客観性を高めるためには評価基準を複数設定することが望ましい。しかし、これは短い評価時間、評価のしやすさなどの実用性と反することになり評価の適切さを損なうことにもなる。この点に関しては、H. E. の方が勝っていると言えるが、信頼性の面では問題点が残る。

これらの評価に関して、内藤(1994)はスピーチの評価点を分析することによって、次の事を明らかにしている。

1) 英語の運用能力がある者が評価者になる場合、SEMI-ANALYTIC EVALUATION と HOLISTIC EVALUATION の間には強い相関がある。

2) 評価は、特に CONTENT と強い相関がある。

また、英語学の専門家は評価点が厳しくなるという松居(1970)の研究があるが、内藤(1994)のデータの中でもそう言えるようだとしている。

(2) WRITING の評価からの示唆 [頁数の関係上省略]

## 3. 実験研究

### (1) 被験者と評価者および評価方法

1) 被験者：高校英語弁論大会の出場者(5番～15番)

[1～4番を除いたのは、評価の信頼性を高めるためである]

2) 評価者：日本人教師6名とALT3名の合計9名

3) 評価方法：A. E. (CONTENT, ORGANIZATION, GRAMMAR, VOCABULARY, BODY LANGUAGE, EYE CONTACT, PRONUNCIATION, TONAL VARIETY の8項目)と S. E. (CONTENT, DELIVERY, PRONUNCIATION の3項目)と H. E. の3種類。3つの評価法とも満点は40で、A. E. は各項目

5 で、S. E. は CONTENT(20), DELIVERY(10), PRONUNCIATION(10) である。  
そして、A. E.、S. E.、H. E. は次のように対応している。

A. E.		S. E.		H. E.	
CONTENT	}	CONTENT	}	}	
ORGANIZATION					
GRAMMAR					
VOCABULARY					
BODY LANGUAGE	}	DELIVERY	}		
EYE CONTACT					
PRONUNCIATION	}	PRONUNCIATION	}		
TONAL VARIETY					

(2) 仮説

- 1) 英語の運用能力がある者が評価者(EVALUATOR)になる場合、  
(A) A. E., S. E., H. E.の間には強い相関がある。  
[内藤(1994)では S. E., H. E.の間には強い相関が見られた。]  
(B) H. E. は主観的と言われるが、かなり信頼性の高い評価方法である。  
(C) 評価は、特に CONTENT と相関が強い。  
(D) 英語学を専門とする評価者は評価点が厳しくなる。

(3) 評価(点)の分析方法

相関関係 [ピアソンの相関係数] (PEARSON PRODUCT MOMENT COEFFICIENT)  
有意差検定 [一元配置法分散分析→ライアンの法] (ONE-WAY ANOVA→RYAN'S METHOD)  
信頼度 [キューダ・リチャードソン 21 の公式] (KUDER-RICHARDSON 21 FORMULA)  
分布 [標準偏差] (STANDARD DEVIATION) 等

(4) 結果および分析

TABLE 1 (9 名の評価者の評価: N=11(以下同様))

	A	B	C	D	E	F	G	H	I
TOTAL	260	331	384	320	314	352	346	337	357
MEAN	23.6	30.1	34.9	29.1	28.5	32.0	31.5	30.6	32.5
SD	2.64	3.32	2.27	3.26	4.46	3.02	4.83	4.21	3.17
MAX	27	35	39	35	34	36	38	38	38
MIN	19	23	31	24	19	27	20	24	28
RANGE	8	12	8	11	15	9	18	14	10

NOTES:

A - C = A. E.  
D - F = S. E.  
G - I = H. E.  
B, C, E, F, H, I  
= JTES  
A, D, G = ALTS

TABLE 2 (TABLE 1 の分散分析およびライアンの法)

\*\*\*\*\* ANOVA \*\*\*\*\*

FACTOR	SS	DF	MS	F
SSB	880.313	8	110.039	7.905
SSW	1252.750	90	13.920	
TOTAL	2133.063	98		

SIGNIFICANCE OF F

\*\*\* p<0.001

\*\*\*\*\* RYAN'S METHOD \*\*\*\*\*

LOCATION OF SIGNIFICANCE

	A	B	C	D	E	F	G	H
B **								
C ***								
D **			*					
E *			**					
F ***								
G ***								
H ***								
I ***								

\*\*\* P<0.001  
\*\* P<0.01  
\* P<0.05

TABLE 3 (9名の評価点の相関)

	A	B	C	D	E	F	G	H	I	
A		.58+	.78**	.81**	.62*	.77**	.66*	.84**	.86***	*** p<0.001
B			.83**	.74**	.56+	.51	.74**	.66*	.77**	** p<0.01
C				.83**	.58+	.57+	.81**	.85***	.77**	* p<0.05
D					.60+	.71*	.85***	.71*	.75**	-----
E						.70*	.85***	.62*	.67*	+ p<0.1
F							.71*	.72*	.86***	
G								.68*	.72*	
H									.90***	

有意相関の数 (21の内 5% 水準(\*)以上の数)

A. E. = 15			S. E. = 15			H. E. = 21													
Cf.																			
***	[	A. E.	2	**	[	A. E.	11	*	[	A. E.	2	+	[	A. E.	4	[	A. E.	1	
		S. E.	3			S. E.	5			S. E.	7			S. E.	4			S. E.	0
		H. E.	6			H. E.	6			H. E.	9			H. E.	0			H. E.	0

TABLE 4 (A. E.、S. E.、H. E. の各々の合計と100点満点への換算[X 0.8333334])

	A. E. (1)		S. E. (2)		H. E. (3)	
	合計	換算	合計	換算	合計	換算
	(120)	(100)	(120)	(100)	(120)	(100)
TOTAL	975	812.4	986	821.6	1040	866.6
MEAN	88.6	73.9	89.6	74.7	94.5	78.8
SD	7.43	6.19	9.47	7.89	11.18	9.33
MAX	100	83.3	102	85.0	113	94.2
MIN	75	62.5	71	59.2	75	62.5
RANGE	25	20.8	31	25.8	38	31.7
RT		0.50		0.70		0.82

TABLE 5 (TABLE 4 [合計] の分散分析およびライアンの法)

***** ANOVA *****					***** RYAN'S METHOD *****	
FACTOR	SS	DF	MS	F	LOCATION OF SIGNIFICANCE	
SSB	219.344	2	109.672	1.108		
SSW	2968.660	30	98.955		1 2	
TOTAL	3188.004	32			2	
SIGNIFICANCE OF F					3	
p< 0.4					NO SIGNIFICANCE	

TABLE 6 (各評価法の相関)

	A. E.	S. E.	H. E.	
A. E		.82**	.92***	*** p<0.001
S. E			.93***	** p<0.01
H. E				* p<0.05

TABLE 7 (A. E. 評価者 3 人の合計)

	1	2	3	4	5	6	7	8	
	CONT.	ORGAN.	GRAM.	VOC.	PRO.	TON.	BOD.	EYE.	計
TOTAL	125	123	130	125	116	128	117	111	975
MEAN	11.4	11.2	11.8	11.4	10.5	11.6	10.6	10.1	88.6
SD	1.37	1.11	0.72	0.48	1.72	1.72	1.77	1.83	7.43
MAX	13	13	13	12	13	14	14	13	100
MIN	9	9	11	11	7	7	8	7	75
RANGE	4	4	2	1	6	7	6	6	25

TABLE 8 (TABLE 7 の分散分析およびライアンの法)

***** ANOVA *****					***** RYAN'S METHOD *****							
FACTOR	SS	DF	MS	F	LOCATION OF SIGNIFICANCE							
SSB	27.885	7	3.984	1.782	1 2 3 4 5 6 7							
SSW	178.820	80	2.235		2							
TOTAL	206.705	87			3							
SIGNIFICANCE OF F					4							
p< 0.2					5							
					6							
					7							
					8							

TABLE 9 (A. E. 各項目の相関)

	CONT	ORGAN	GRAM	VOC	PRO	TON	BOD	EYE	計	
CONT		.14	.35	.49	.49	.37	.02	.24	.53	*** p<0.001
ORGAN			.27	.22	.56+	.56+	.22	.66*	.69*	** p<0.01
GRAM				.46	.30	.17	.31	.29	.48	* p<0.05
VOC					-.02	-.06	.37	.38	.39	-----
PRO						.89**	-.02	.56+	.77**	+ p<0.1
TON							.25	.67*	.83**	
BOD								.71*	.56+	
EYE									.90***	

TABLE 10 (S. E. 評価者 3 人の合計 : SD 比較のため DEL PRO の各項目 20 点満点に換算 : 3 人の合計は 60 点満点)

	CONT(1)	DEL(2)	PRO(3)	合計
TOTAL	497	474	438	1409
MEAN	45.2	43.1	39.8	128.1
SD	5.52	7.15	7.36	17.76
MAX	52	54	48	148
MIN	35	30	24	89
RANGE	17	24	24	59

TABLE 11 (TABLE 10 の分散分析およびライアンの法)

***** ANOVA *****					***** RYAN'S METHOD *****							
FACTOR	SS	DF	MS	F	LOCATION OF SIGNIFICANCE							
SSB	163.012	2	81.506	1.637	1 2							
SSW	1493.390	30	49.780		2							
TOTAL	1656.402	32			3							
SIGNIFICANCE OF F					p< 0.3							
					NO SIGNIFICANCE							

TABLE 12 (S. E. 各項目の相関)

	CONT	DEL	PRO	合計	
CONT		.90***	.85***	.98***	*** p<0.001
DEL			.79**	.94***	** p<0.01
PRO				.92***	* p<0.05
					-----
					+ p<0.1

TABLE 13 (A. E. を S. E. に対応させた各項目の相関)

	CONT	DEL	PRO	合計	
CONT		.55+	.45	.79**	*** p<0.001
DEL			.41	.82**	** p<0.01
PRO				.79**	* p<0.05
					-----
					+ p<0.1

## (5) 考察

9名の評価者について、標準偏差は 2.27 から 4.83 と分布に幅が見られる。最高点と最低点の幅についても同様である。(TABLE 1) 分散分析によれば、9人の評価点の平均には 0.1% 水準で有意差があり、ライアンの法では A が他のすべての評価者と 5% から 0.1% 水準で有意差がある。また、C も D、E と 5% から 1% 水準で有意差がある。(TABLE 2) すなわち、A は評価点が最も小さい 23.6 で C は最も大きい 34.9 である。従って、A. E. 内に点差が大きいことを示している。

そして、評価点が最も厳しいのが A の 23.6 で、続いて E の 28.5 である。英語学の専門家は評価点が厳しくなる傾向があると言われているが、この場合 A も E も英語学の専門である。さらに、内藤(1994)のデータでも全く同様の結果が得られている。[1994 年のデータ: 27.5 から 34.1 の内 27.5 に続いて 28.5 が 2 名とも英語学の専門家] 従って、仮説 (D) はほぼ証明されたことになる。松居(1970)も述べているように、英語学の専門家は評価点が厳しくなる傾向があると言えよう。

さて、9名の評価点の間には強い相関がかなり見られる。有意な相関 (5% 水準以上) の数については、A. E. が 15、S. E. が 15、H. E. が 21 と、これらの中では H. E. が一番多い。(TABLE 3) また、A. E.、S. E.、H. E. を合計し、それらを KR-21 で信頼度係数を求めるため 100 点に換算すると平均点では A. E. が 73.9、S. E. が 74.7、H. E. が 78.8 である。標準偏差は A. E. が 6.19、S. E. が 7.89、H. E. が 9.33 となっている。これらの信頼度係数は、A. E. が 0.50、S. E. が 0.70、H. E. が 0.82 で H. E. が最も高く、続いて S. E.、A. E. となっている。(TABLE 4) 分散分析の結果では、これら 3つの評価法の間には有意差はない。(TABLE 5) そして、これらの評価法の間には、A. E. と S. E. に 0.82\*\*、A. E. と H. E. に 0.92\*\*\*、S. E. と H. E. に 0.93\*\*\*と非常に強い相関がある。(TABLE 6) 従って、仮説 (A) は証明されたことになる。さらに、特に H. E. と他の評価の相関は強く、9名の評価者の中の有意相関の数でも H. E. が一番多く、信頼度係数も 0.82 と最も大きいので、H. E. の信頼性は 3つの評価法の中では最も高いと考えられる。従って、仮説 (B) も証明されたことになる。

A. E. 3名の評価者の各項目を見ると、平均は EYE CONTACT の 10.1 から GRAMMAR の 11.8 までである。標準偏差は VOCABULARY の 0.48 から EYE CONTACT の 1.83 までである。(TABLE 7) 標準偏差によれば発音や表現力において分布に幅があると言えよう。しかし、点数においては、分散分析の結果それぞれの項目間に有意差はない。(TABLE 8) A. E. 各項目の相関は、それぞれあまり強いとは言えないが、TONAL VARIETY と PRONUNCIATION が 0.89\*\*、EYE CONTACT は BODY LANGUAGE と 0.71\*、ORGANIZATION と 0.66\*、TONAL VARIETY と 0.67\* と有意な相関がある。そして、合計と有意な相関があるのは EYE CONTACT 0.90\*\*、TONAL VARIETY 0.83\*\*、PRONUNCIATION 0.77\*\*、ORGANIZATION 0.69\* である。(TABLE 9) 発音や表現力は合計との相関が強いと言えよう。

S. E. 3名の評価者の各項目について、全てを 20 点満点にした平均点は CONTENT が 45.2、DELIVERY が 43.1、PRONUNCIATION が 39.8 である。標準偏差は CONTENT が 5.52、DELIVERY が 7.15、PRONUNCIATION が 7.36 である。A. E. と同様、発音や表現力において分布に幅があると言え

よう。(TABLE 10) しかし、分散分析の結果、点数において有意差は見られない。(TABLE 11)

S.E. の3つの項目間には、それぞれ有意な強い相関が見られる。特に、評価の合計と CONTENT は 0.98\*\*\*と非常に強い相関がある。(TABLE 12) しかし、A.E. を S.E. の各項目にまとめて対応させたものでは、各項目間に有意な相関は見られない。(TABLE 13) そして、評価の合計と DELIVERY が 0.82\*\* の相関で最も強く、CONTENT とは 0.79\*\*である。従って、S.E. においては、評価は特に CONTENT と相関が強いと言えるが、A.E. においては、そうとは言えない。従って、仮説(C)は S.E. においては支持されるが、A.E. においては支持されたことにはならない。また、内藤(1994)のデータでも S.E. においては同様の結果が得られている。

#### 4. おわりに

A.E. は、被験者への FEEDBACK がしやすいという長所がある。すなわち、どの項目において優れ、どの項目において劣っているかなどを明示することができる。また、より客観的に評価することもできる。しかし、項目が多く、評価にも時間がかかり煩雑である。煩雑な故に評価に歪みが生ずる可能性がある。今回、実際に評価をし、データを分析してみると、音声言語である SPEECH の評価には、分析的な評価であっても S.E. の方がより妥当であると思われる。これは、データ分析のみならず、アンケートによっても A.E. の評価者全員が分析的評価は S.E. の方がより妥当であろうとしている。また、H.E. も今回の分析では信頼性が高く、特に実用性においては非常に優れた評価方法と言えよう。

(福井県立 鯖江高等学校)

#### 引用文献

- 福井 保 1969. 「英語テストとその形式(5)音声テストー話すことのテスト」  
『現代英語教育』8月号 p.14. 研究社出版
- 松居 司 1970. 「Oral Production のテスト問題作成と評価という観点から」  
『現代英語教育2月号』 p.17. 研究社出版
- 内藤 徹 1994. 「分析的評価と全体的評価ー高校英語弁論のデータよりー」  
『中部地区英語教育学会 紀要24』 pp.199-204.
- 小川芳男 1982. 『英語教授法辞典』 pp.595-598. 三省堂
- Valette, R.M. 1977. *Modern Language Testing*. 2nd Ed. pp.157-161. New York: Harcourt Brace Jovanovich, Inc.
- Yorozuya, Ryuichi and J.W. Oller, Jr. 1980. "Oral Proficiency Scales: Construct Validity and the Halo Effect," *Language Learning*, 30,1 pp.135-153.